HOW TO GET OUTSOURCING RIGHT FOR DATA ANNOTATION SERVICES?

Abstract

Data annotation is a prerequisite in ML. It involves labelling data to train AI- and ML-based algorithms. To understand the model's use case, it must be trained on good quality and highly relevant data. Hence, optimising the process with accuracy, quality, speed, and security is crucial. The annotation process may demand more than just the knowledge of the domain. An experience and thorough understanding of the AI/ML model in use is an added advantage. Outsourcing can bring the required skills to the annotation process. In this POV, we will explore the ways to choose the right outsourcing partners.

Infosys®
Navigate your next

Challenges with data annotation

The pandemic accelerated the demand for technology-led experiences that heavily rely on AI and machine learning (ML) for their success. As a result, the need to process data rapidly and accurately has increased. The latest round-up on ML's influence suggests that the global ML market will reach $30.6 billion by 2024. Presently, 1 in 10 businesses now use 10 or more AI applications.[1] Developing an ML-based or an AI-based model that can make decisions like humans requires accurate training data to help the algorithms understand the context.

Self-driving vehicles, robots in healthcare and agriculture industries, and many other verticals are garnering upward investment, which drives the need for AI. Behind the success of every AI application is the unseen data annotation service that is crucial for its effectiveness in improving business and customer insights. Labelling data is a crucial process, and errors can mar the performance of the AI model, rendering it useless. Several internal and external issues make data annotation a challenge.[3][4]

- Since data annotation is a continuous process, quality and consistency are of utmost importance for the AI model to function correctly. But, since the field is comparatively new, there's a lack of skilled workforce to deliver top-notch quality reliably.

- As data streams widen steadily, businesses need larger teams to handle the data annotation process. However, managing a large workforce can be expensive and lead to organisational difficulties.

- Along with a skilled workforce, AI/ML models demand cutting-edge tools and technologies to train the data accurately.* But poor process knowledge and lack of infrastructure deter businesses from accessing the latest annotation tools and technologies.

- Labelling data for annotation is a lengthy process and requires businesses to invest in office space, talent, and expensive technology. These factors adversely impact finances due to massive investment required in infrastructure for data annotation.

- Data collected for annotation can include highly personal information. To safeguard the interests of data sources, businesses across the globe are implementing stringent data privacy laws. Failing to comply with data security and privacy standards due to poor process knowledge can lead to data leaks, which affect business reputation.

Types of annotation services

Accurate training data is the key to the success of an AI/ML model for it to deliver value on a large scale. Annotation may not seem complicated, but it requires a qualified and skilled team of data professionals. This allows businesses to manage the quantities of data and deal with bigger projects. Depending on what companies choose, data annotation can be in-house or outsourced to a third party.

In-house data annotation

The need for greater control, data security, and privacy drives businesses to have an in-house labelling team.[5] Moreover, if the ML project is for a long term with the intent to develop a continuous dataset, businesses can opt for an in-house team for data annotation. Some of the benefits of an in-house annotation strategy are:

- Consistent annotation provides significant advantage due to better supervision of an in-house team. Also, the error rate is lower for in-house data teams.

- Close cooperation across verticals and the development team ensures better accuracy and quality.

- In-house data annotation gives the team a better understanding of the task

as data professionals have enough time to learn more about the project.

- There is greater flexibility to alter instructions in an in-house team.

- With greater flexibility, it's faster to deploy changes in ML models, which improves the time to market.

However, huge costs of managing a large workforce and infrastructure, scalability, and skill gaps affect the quality of in-house data annotation. Hence, more and more businesses now prefer outsourcing data annotation services.

Outsourcing data annotation

Outsourcing is typically done through a third-party data annotation service provider or crowdsourced data professionals. Because of lower costs, crowdsourced data annotation has become popular. It is only suitable when annotation needs do not involve sensitive data and the project is for a short term. Moreover, crowdsourcing may not deliver the same quality and consistency as a dedicated annotation service provider. Choosing a third party for outsourcing data annotation is a cost-effective, scalable solution to maximise the accuracy of AI and ML models.[6]

- Deep expertise with a strong talent pool leads to fewer errors. Specialist providers have domain-specific workflows and tested operational processes required for industry-compliant data management.

- Outsourcing maintains consistent data quality. Moreover, competent service providers use state-of-the-art tools and software that allow businesses to review and monitor their progress.

- A dedicated team provides scalability solutions allowing businesses to implement changes as needed by the ML model. Additionally, customised solutions help businesses satisfy different ML projects' needs.

- Any bias can alter how AI/ML models understand the training data to make decisions. Outsourced data annotation helps mitigate internal biases to improve the accuracy of AI and ML models.

- Data security is built into the process to avoid leaks and privacy breaches. The secure annotation process is backed by a business continuity plan to manage contingencies.

How to choose the right annotation service provider?

When you outsource the data annotation work to an external provider, you expect results to meet expectations. Look for these pointers while choosing the right partner for your AI/ML model.

- **Expertise:** The outsourcing partner must have a dedicated pool of experts for error-free data training. A competent annotation technology partner will help businesses through the beginning stages of taxonomy creation and labelling specification to provide a smooth experience later on. While crowdsourcing might be the cheaper option, it suffers from inconsistencies in the skill sets of annotators, making it a choice only if your project is relatively simple.[7]

- **Adaptability:** The annotation service provider should be able to adapt and innovate in a dynamic environment with proven results, which ensure agility for the business. The outsourcing partner should be up-to-date with the rapid changes in the industry and be willing to innovate in unique situations.[7]

- **Technically progressive:** Access to leading-edge technology to work in a platform-agnostic manner helps businesses implement processes quickly into the existing environment. A service provider with the latest tools and technology for annotation can deal with new challenges by customising features in the software for the best results. It indicates a deep understanding of the process and how AI will use that data.[7]

- **Security:** An ideal annotation service provider will place data security as a priority. Check if the outsourcing partner has any international certifications or accreditations. Moreover, the partner must abide by the latest regulations in different regions. A stringent QA framework assures that data leaks are avoided at all times. However, in the event of a breach, check if there is a business continuity plan to manage contingencies.

- **Speed:** By implementing time-efficient processes, an annotation service provider should improve speed to market. Data annotation can be a time-consuming activity, but service providers must have the capacity to meet deadlines while maintaining quality.

Here are a few tips to consider while choosing an annotation service provider:

- Know what type of data you will be working with and the type of annotations required.

- Clearly define the requirements and objectives of the project.

- Perform a quality check by asking to annotate a sample task.

- Examine security to ensure compliance.

- Check pricing and budget.

Conclusion

Supply chain businesses that understand the apparent technological evolution are already investing in AI-based optimisation and transformation. According to a study, 61% of supply chain executives reported that AI helped lower the cost of supply chain management.[3] If implemented successfully, AI reduces costs, increases revenue, improves productivity, and reduces errors and mishaps. This makes AI imperative for all businesses that handle complex supply chains.

*For organisations on the digital transformation journey, agility is key in responding to a rapidly changing technology and business landscape. Now more than ever, it is crucial to deliver and exceed organisational expectations with a robust digital mindset backed by innovation. Enabling businesses to sense, learn, respond, and evolve like living organisms will be imperative for business excellence. A comprehensive yet modular suite of services is doing precisely that. Equipping organisations with intuitive decision-making automatically at scale, actionable insights based on real-time solutions, anytime/anywhere experience, and in-depth data visibility across functions leading to hyper-productivity, Live Enterprise is building connected organisations that are innovating collaboratively for the future.

References

1. https://www.forbes.com/sites/louiscolumbus/2020/01/19/roundup-of-machine-learning-forecasts-and-market-estimates-2020/?sh=35260e3f5c02

2. https://datafloq.com/read/outsource-data-annotation-services-to-accelerate-product-development/

3. https://maxicus.com/challenges-in-data-annotation/

4. https://www.linkedin.com/pulse/why-better-outsource-data-annotation-suntecindia?trk=pulse-article_more-articles_related-content-card

5. https://labelyourdata.com/articles/choose-data-annotation-in-house-outsourcing-platform#:~:text=move%20to%20make.-,%E2%80%94%20Data%20Annotation%20Outsourcing,-Why%20should%20companies

6. https://appen.com/blog/5-reasons-to-outsource-your-data-annotation-projects/#:~:text=A%20huge%20advantage%20of%20outsourcing,than%20most%20internally%20resourced%20teams.&text=This%20helps%20ensure%20the%20highest%20level%20of%20data%20quality.

7. https://insidebigdata.com/2020/02/16/why-partnerships-beat-outsourcing-in-data-labeling/

For more information, contact infosysbpm@infosys.com

Infosys®
Navigate your next

Infosysbpm.com

Stay Connected